Methodology for the derivation of the INCGROSS variable, using administrative sources

Astrid Depickere

STAT**BEL**

Outline

- Context

- Available Sources

- Coverage of administrative data

- Methodology

- Some results

- Evaluation

STAT**BEL**

Context

- New definition of INCGROSS in IESS FR

- Extended transmission deadline (Y + 15 months)

- Improved quality of administrative databases

- Improved procedures and delivery times

- Large pressure on administrative simplication ("Only once" law)

- Grant project 2019-2020 : feasability study

- Reformed questionnaire from 2021 on: no more question on income from work

STAT**BEL**

Available sources

Social Security data: DMFA (Déclaration multifonctionnelle/multifunctionele Aangifte)

- Quarterly basis: declaration of employers to social security administration

- Detailed information on wages and working times

- Availability: 1st version Q+5 months, final version Q+12 months

Tax register data

- Provisional income tax data (Belcotax): available Y+6 months

- IPCAL: more complete version of personal income tax data: available Y+16 months

   => after transmission deadline for INCGROSS

Coverage of administrative data

Table 1: Missing values after linking LFS 2020 data to administrative sources of Social Security (RSZ) and Tax register (BTAX)

		RSZ/DMFA		BTAX		BTAX & RSZ/DMFA	
		missing	not missing	missing	not missing	missing	not missing
Student (mainstat)	N	42	175	23	194	23	194
	Row Pct	19,4	80,7	10,6	89,4	10,6	89,4
Disabled (mainstat)	N	185	61	77	169	77	169
	Row Pct	75,2	24,8	31,3	68,7	31,3	68,7
Working abroad	N	666	53	668	51	666	53
	Row Pct	92,6	7,4	92,9	7,1	92,6	7,4
NACE U (extraterritorial organisations and bodies)	N	184	31	176	39	125	77
	Row Pct	85,6	14,4	81,9	18,1	61,9	38,1
Temporary job	N	190	1120	139	1171	120	1190
	Row Pct	14,5	85,5	10,61	89,39	9,16	90,84
Absent from work	N	603	1876	272	2207	268	2211
	Row Pct	24,3	75,7	11,0	89,0	10,8	89,2
All	N	1.659	12.656	1.215	13.100	1.172	13.143
	Row Pct	11,6	88,4	8,5	91,5	8,2	91,8

STATBEL
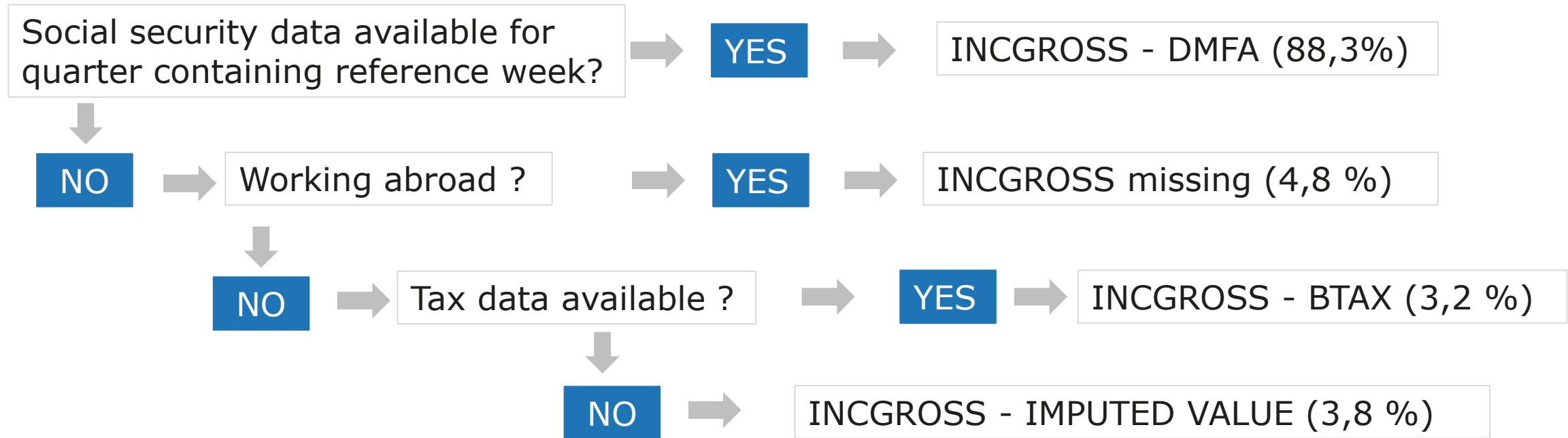
Methodology – Social Security data

Step 1: select employment line(s) corresponding to LFS main job in reference week, different scenario's :

- Only one employment line + period to which it applies contains the reference week => assumption this is the main job

- Multiple lines, no overlap between lines => select the line that contains the reference week

- Multiple lines that contain the reference week (overlap) => assumption that the line with the highest salary applies to the main job

- None of the employment lines contain the reference week => select the employment line closest to the reference week

Step 2: recalculate the salary to a monthly basis (SES concept)

- Full time workers, using the number of working days

- Part-time workers, using the working hours

STAT**BEL**

Methodology – Belcotax data

- Mainly for those absent from work (illness, disability, temporary unemployment, parental leave,…), temporary jobs, student jobs,…

- Only if no information is available in social security database

- Not used for persons working abroad

- Currently also used for those working in NACE U (extraterritorial organisations & bodies)

- Derivation of Gross Monthly Income from Work (-> SILC concept)

- Drawbacks:

  - Calculated income from work can apply to multiple jobs

  - Yearly basis -> monthly basis

STATBEL Some results: LFS vs SES

Table 3: Comparison gross monthly wage in LFS – SES 2020 (reduced sample, full time working employees)

Analysis Variable : INCGROSS_				SES			Diff SES vs INCGROSS	
(NACE Rev. 2, 1 digit)	N Obs	Mean	Median	N	Mean	Median	Mean	Median
Mining and Quarrying	13	4.610	3.339	220	4.367	3.900	106%	86%
Manufacturing	1.404	3.988	3.293	25.930	3.905	3.637	102%	91%
Electricity, Gas, Steam and Air Conditioning Supply	77	5.538	5.670	896	5.316	5.179	104%	109%
Water Supply; Sewerage, Waste Management and Remediation Activities	93	3.430	3.210	1.186	3.926	3.706	87%	87%
Construction	505	3.170	2.904	7.882	3.372	3.293	94%	88%
Wholesale and Retail Trade; Repair of Motor Vehicles and Motorcycles	801	3.278	2.846	20.220	3.643	3.354	90%	85%
Transportation and Storage	603	3.203	2.892	10.287	3.437	3.195	93%	91%
Accommodation and Food Service Activities	138	2.479	2.371	1.374	2.789	2.789	89%	85%
Information and Communication	347	4.322	3.965	5.487	4.620	4.395	94%	90%
Financial and Insurance Activities	307	5.304	4.722	4.353	5.071	4.874	105%	97%
Real Estate Activities	35	4.002	3.469	515	4.501	4.116	89%	84%
Professional, Scientific and Technical Activities	311	3.855	3.411	6.855	5.260	4.909	73%	69%
Administrative and Support Service Activities	401	2.713	2.601	17.234	3.348	3.228	81%	81%
Size of local unit	N Obs	Mean	Median	N	Mean	Median	Mean	Median
10-19	640	3.028	2.791	6.239	3.426	3.239	88%	86%
20-49	980	3.244	2.902	17.216	3.576	3.361	91%	86%
50-249	1.843	3.428	3.008	41.470	3.793	3.551	90%	85%
250-499	541	4.079	3.465	19.011	4.015	3.671	102%	94%
500 +	1.031	4.652	3.877	18.503	4.401	4.083	106%	95%

STATBEL Some results: INCDECIL vs INCGROSS

Table 4. Comparison of old and new LFS wage variable, using a categorical variable (Quintiles)

QUINTILE_INCGROSS	QUINTILE_q100def 1	2	3	4	5	Total
1	**1.741**	482	322	157	117	2.819
2	633	**1.139**	581	220	154	2.727
3	141	709	**1.136**	417	252	2.655
4	69	229	823	**1.081**	463	2.665
5	20	56	168	650	**1.889**	2.783
	2.604	2.615	3.030	2.525	2.875	13.649

Frequency Missing = 666

Same quintile	**6.986**	**51%**
Adjacent quintile, INCGROSS > Q100def	2.815	21%
Adjacent quintile, INCGROSS < Q100def	1.943	14%
More than 1 quintile difference, INCGROSS > Q100def	683	5%
More than 1 quintile difference, INCGROSS < Q100def	1.222	9%
	13.649	100%

STAT**BEL**

Conclusion

New approach can be considered as an improvement compared to the old approach:

- Reduction of response burden

- More accurate measurement

- Reduction of number of missing values

Some remaining problem areas:

- Employees working abroad

- Employees working for extraterritorial organisations and bodies

Re-introducing a question for specific groups may be necessary